

(Intro Music Starts)

Sadie Witkowski: Hello, I'm Sadie Witkowski.

Ian Martin: And I'm Ian Martin.

SW: And you're listening to Carry the Two, a podcast from the Institute of Mathematical and Statistical Innovation, or IMSI.

IM: We talk about the real-world applications and implications of mathematical and statistical research.

SW: And this season, we're focused on how artificial intelligence can aid, and sometimes hurt, decision-making.

IM: People are trying to use AI for everything from cutting-edge research down to relationship advice or dinner recs. We want to understand what the mathematics of AI and decision-making can tell us about how AI will impact these choices.

SW: And the ifs and wheres of using it at all.

(Intro Ends)

IM: So Sadie, last time we chatted with Jeremy Avigad at Carnegie Mellon University, and he told us about how mathematicians are using different types of AI in their research.

SW: That's right.

IM: And he talked a lot about how they're trying to use AI to verify that the math is "right."

SW: Yup!

IM: Well, that was very helpful for understanding how people are thinking about truth, but I'm also curious about the decision-making part.

SW: Makes sense. It is part of the season's title, after all.

IM: I mean, we just talked about it in the intro!

SW: Well, you're in luck. This time, we're speaking with someone about (emphasized) *exactly* that. "Decision-making" is in the first line of his paper, even!

IM: Ooh, who is that?

SW: I'll let him introduce himself.

Saptarshi Pal: Sam Saptarshi Pal, you can refer to me as Saptarshi. I am a postdoctoral researcher [and] lecturer at Harvard right now in the department of mathematics.

SW: Saptarshi works on game theory and cooperation in games. And recently, he, along with a group of researchers from Harvard, Dartmouth, and the University of Munich, [took a look](#) at how AIs and specifically large language models approach games.

IM: Cooperative games? I love [Hanabi](#)!

SW: (sarcastic) Ha ha, I know you know what game theory actually is.

IM: Fine. I love the prisoner's dilemma!

SW: That's more like it. Why don't we take a little refresher on the prisoner's dilemma, since that's actually what they looked at.

IM: Nailed it. (beat) So, the prisoner's dilemma supposes that you and another player are stuck in jail, you're separated and can't talk to each other, and you're both given an offer by the police. They want each of you to sell the other out. If both of you stay silent, you both get off with a light sentence. If only you, or only your partner, talk, then whoever betrayed gets no jail time while the other gets a bad sentence. If you both sell the other out, though, you both get sentenced, more than if you both stayed silent but less than if only one of you betrayed the other.

SW: Exactly, and they looked at the "iterated prisoner's dilemma," specifically. So you both choose whether to betray or not, then you both see the result, and then you get re-sentenced and have to do the whole song and dance all over again.

IM: That gets complicated fast.

SW: For one round, you have two options—stay silent or betray—but if you think about two rounds you essentially have 4 options—do either twice, cooperate then betray, or vice versa. In three rounds you have 8, then 16, and on and on.

SP: Prisoner's dilemma is a great model for cooperation because it's so simple, but at the same time it's so rich, mathematically speaking, especially the repeated version of it...The complexity increases depending on how far back you remember.

IM: If you're at the point where you are playing the repeated prisoner's dilemma with someone, maybe, just maybe, you two should stop going on crime sprees together.

SW: I'm innocent! I'll never talk!

IM: (laughs) So, how does this relate to LLMs, then?

SW: They wanted to know if all the training and tuning that models go through end up imbuing the models with strategy. See, if lawyers and policymakers and all sorts of people are going to prompt LLMs when they have to make high-stakes decisions, then the least we could hope for is that these bots are giving rational advice.

IM: Even more so, some people want to set these AI models out on the web to make decisions on their own.

SW: Right! Those are called agents! So, Saptarshi wanted to understand the *strategy* behind what choices are made in the repeated prisoner's dilemma.

SP: A strategy is kind of a lookup book in your pocket that you can just look up, okay this is the history that occurred, my book is telling me to play 'cooperate' after this particular history.

IM: So whether they betray or not based on everything that already happened.

SW: Like a flowchart!

IM: A really big flowchart, if you play long enough.

SW: In particular, Saptarshi and his co-authors wanted to know whether an AI's strategies are something called Nash equilibrium or not.

IM: Nash, as in Music City Nashville, Tennessee?

SW: No, Nash as in John Nash, the mathematician!

IM: Sounds familiar...

SW: *A Beautiful Mind*?

IM: Right, right. I totally remembered that. (giggles) While we're at it, though, can you remind our listeners what a Nash equilibrium is? I mean, I remember, but for their sake.

SW: (sighs)

SP: A Nash equilibrium strategy is a strategy which, if both players play the strategy, there is no payoff incentive for any one player to deviate from that strategy.

IM: Wait, okay, that's a little hard to keep in my head at once.

SP: If I am playing a strategy A, and you're playing the same strategy A, we would call this strategy a Nash equilibrium strategy if you have no payoff incentive to switch to any other strategy B given I keep playing A.

SW: You can think of this in a single prisoner's dilemma pretty easily. Staying silent is not Nash equilibrium because the other player could betray to decrease how much time they serve. Defecting is Nash equilibrium because the other player can't do any better than also defecting.

IM: They do worse if they don't defect, even!

SW: Exactly. They also looked at if the strategies were "partners" or "rivals."

IM: That's how I approach most of my relationships.

SW: (mock offended) Excuse me? What am I, then?

IM: Wouldn't you like to know. What's a rival then, partner?

SW: ...A rival always does at least as good as its opponent no matter what.

IM: And a partner, rival?

SW: (sighs) They're just a type of Nash equilibrium strategy where both sides always work together if they both use it.

IM: Got it! So, they just had the AIs "play" the game then?

SW: Not exactly. Remember how we said there's exponentially many decisions to make for each extra round you do? This could overwhelm compute capacity quickly, so they realized they could just measure each round one-by-one by feeding it all the possible previous rounds. Basically, they made up a gameplay history and fed that into the LLM as part of the prompt

SP: We don't actually make the LLMs play real games against agents or humans, so they're not playing the games in real time. We just say: "In the past round, you cooperated and the other player defected. This is a new round. What would you like to do?"

SW: In the paper, they call those "memory-1" strategies. They also looked at how they did when looking two rounds back. Then they just keep seeing what the models do a bunch of times!

SP: We go over these histories, we ask it, 'what would you like to do after this particular history,' and then we see whether it cooperates or defects, and we note it down over 50 times. By doing this, we have an estimate for its strategy.

IM: 50?! I can see why they didn't do the whole game, then, that's a lot!

SW: I KNOW!

IM: So, what did they find? What strategies are happening, if any?

SW: Funny you should say this, Saptarshi did admit one thing is for sure.

SP: They seem to be overly cooperative or stupid, I don't know which one it is.

IM: Well, (pause) that's relieving. (laughs)

SW: (laughs) Not exactly the robot overlords you hear about online, huh.

IM: For sure. What do these cooperative strategies look like?

SW: Have you heard of tit-for-tat?

IM: I can take a stab at it. If you don't sell me out, I won't sell you out next time. But if you do, I'm coming for you, next.

SW: Exactly. So that's cooperative, and a partner strategy. The LLMs tended to do strategies like this. For example, they witnessed Anthropic's Claude model prefer something called the Forgiver strategy.

SP: So, the Forgiver strategy is very simple. In the first round it will cooperate. It will cooperate after both players have cooperated, so it's a mutually cooperative strategy. It does not cooperate if, in the previous round, I cooperate and then you defected, but it will cooperate if in the last round I defected and you cooperate or if both of us defected. So it cooperates in every situation except in the case where I cooperated and you cheated me, so it's a very forgiving strategy.

IM: Hence the name Forgiver.

SW: Very perceptive. Congrats, you are at least memory-1.

IM: I try.

SW: Weirdly, it sounds like this is not a very popular strategy among game theorists, so why Claude acts as such is anyone's guess.

SP: Theoretically, it has some good properties, and there has been some research on it in the past, but it has not been as popular as tit-for-tat. So, we were also very surprised to find out that Claude's model was coming up with the Forgiver strategy.

SW: Now, is the LLM “aware” of the Forgiver strategy, or was it just arbitrarily picking in a way that mirrored the strategy, they couldn’t test with their methods, but it is odd.

SP: When we did our experiments, we never told the LLMs to give me a strategy if this is your game. If that were the case, I could imagine that an LLM actually sits down and does the math and goes over the literature, reasons it, and gives me a strategy. All the latest models can do this, and give me a Nash equilibrium strategy that is also cooperative, but rather we actually asked it over all the histories independently.

SW: Note that reasoning models, which Saptarshi is referring to, are a specific type of model that are trained specifically to generate many steps to “walk through” a response to a prompt. These are the ones that are making a bunch of waves in math since they can in theory work a few steps ahead, like he’s talking about. OpenAI’s GPT-5, which they tested, has a reasoning model, for example.

IM: Huh. So, the Forgiver strategy is Nash equilibrium, then?

SW: (elongated) Well, that’s where “it depends” is going to start striking.

IM: Those dreaded words.

SW: We’ll talk about all that right after this commercial break.

(Ad Break)

(End Ad Break)

SW: So, before the break we were talking about how Saptarshi evaluated AI strategies in the prisoners dilemma. They did a whole bunch of different tests, across 5 different models. The ones we’ve been talking about, some that are more actually similar to two players playing, some prompted ones—which I cannot WAIT to tell you about—and the iterated games both for finite and infinite simulations.

IM: Oh, wow. Must have burned through a lot of tokens doing all that.

SW: For sure. They also put all the prompts they used in the paper, so if anyone wants to try it for themselves, they can!

IM: But the “it depends”!

SW: Right! So, these cooperative strategies can be Nash equilibrium when the prisoner’s dilemma is unbounded, but *not* when it has a fixed length.

IM: So, like if it’s 10 rounds.

SW: Yes, or even 1. Remember when I said earlier that betraying is the only Nash equilibrium strategy for the regular, one-shot prisoner's dilemma?

IM: No, sorry, I'm only memory-1 unfortunately.

SW: (grumbles) Well, I did. (regular) And you can show inductively that that means betraying all the time is the only Nash equilibrium strategy for any finite number of games, provided everyone knows how many rounds they'll play at the start.

IM: Oh, so this means that all of these cooperative strategies aren't optimal there.

SW: Right. So when the game goes on infinitely, and the LLMs are prompted with this, all of them choose cooperative strategies that are all partner strategies.

IM: —and so Nash equilibrium.

SW: And if the games are long but could stop, four of them try to cooperate and not sell out their partner but only two are Nash equilibrium about it.

IM: Which two, just so I know which are “more strategic than a 5th grader” or however you want to phrase it?

SW: OpenAI's GPT-4o and Meta's Llama, for the listeners' reference.

IM: Got it. And what about when it's finite, then?

SW: That's where they're all overly cooperative, as Saptarshi put it. He was surprised that only one of them even betrayed on the last round.

SP: This creates an incentive for players to not cooperate in the last round. So we wanted to test this fact, so are LLMs aware of this, so they knew that they're going to only play 10 rounds.

IM: So, betraying your fellow prisoner is just free in the last round, there's no consequences.

SP: We actually found that all of our five LLMs, except GPT-5, did not do this. They cooperate in all of the rounds despite knowing that there are finitely many rounds to play. GPT-5 just cooperated in all the 9 rounds, but in the last round it defected.

SW: Hence, overly cooperative.

IM: Okay, okay, maybe none of them are exactly more strategic than a 5th grader, or maybe a 5th grader wouldn't know to do that either. So, why do they do this? Do they know?

SW: Not exactly.

SP: It would be nice to know where that comes from. I mean, it's all in the training, and it's all in what is called RLHF—I think it's called reinforcement learning with human feedback—so I understand vaguely where their overly cooperativeness comes from.

SW: But they couldn't say exactly. RLHF, as Saptarshi pointed out, is used by all the big AI companies to make their models more aligned. Among other things, it involves humans manually grading AI-generated responses based on how “helpful” or “safe” or “nice” they are, which is then used to further train the model to output responses that are more these things in the future.

IM: So that all explains the HF.

SW: The “RL” or reinforcement learning is the whole last bit, where the human grades helped train a model that can give “reward” signals to the original one. That idea's fairly old in AI research, but the RLHF all together is much newer. Companies are even trying to replace the human feedback step with more AI to make it cheaper, now.

IM: Huh, okay. So, maybe the signals for ideas like “helpful” or “safe” are similar to cooperativeness, then?

SW: Maybe, that's one idea! But again, Saptarshi and his co-authors couldn't draw any conclusive evidence yet.

SP: The biggest takeaway is that LLMs still need...like, they're still not perfect at repeated games. We find that some of the time they're doing very good, but in other times they are...they're falling short.

SW: Saptarshi says this could show how LLMs are progressing in strategy as they get better, or it could expose how smarter players—be they humans or other AI models—could exploit them.

SP: They reason pretty well, but their reasoning is not entirely accurate, and that makes them interesting agents.

IM: Huh, fascinating. Well thanks for telling me all this. Anyway, let's wrap it up.

(outro music)

IM: Don't forget to.

SW: WAIT!

(outro music abruptly stops)

SW: I NEEDED to tell you about the whole persona part!

IM: Persona part?

SW: Where they prompt the models to play the game as different personas.

IM: Oh, I think my little memory-1 memory is kicking in. You were saying this was another variant.

SW: Yes! And I think it is also an important point to leave our listeners with!

IM: Okay, so what does this all mean?

SW: Right, so they *also* found that how they phrased all the information mattered a ton as to how the models responded. Bring Saptarshi back in!

SP: We saw that the behavior, or rather the effective strategies of these models, are very susceptible to these prompts. In particular, if we had this sentence where we added some objective.

IM: And these are the personas?

SW: Yes, things like “get as many points as you can” or “beat the other player.” But my favorite one was “play like a pro.”

SP: That was the idea of the most senior author in the paper.

SW: He calls it the “Playoff Pro.”

SP: I think Playoff Pro is interesting because it’s kind of like, if you treat an LLM as a very, very smart person—which it’s not—they might be confused. You know, they might be thinking, “Oh, we are already playing like a pro, like what do you mean, like should I do something different now?”

SW: So this persona actually didn’t do anything, or even misled the model outputs sometimes!

SP: We did see that most of the LLMs kept playing what they were playing, but some of them made some changes. I think the point of that was to create some ambiguity. I mean, it’s not really an objective to play like a pro. It’s like, it’s not even clear what it means.

SW: And this kills me! I’ve definitely seen some people put that kind of language in their prompts when using AIs.

IM: I can believe it, gaming the system and all that.

SW: Exactly! And it's like, people think that's really clever, but it doesn't even necessarily make sense!

SP: In that sense, we show that, like, it makes no difference, like if you add sentences like that, it just keeps doing the same thing.

SW: So THERE'S your other takeaway, dear listeners. Don't bother trying to imbue your chatbot with some abstract cleverness—it doesn't work!

IM: (chuckles) You know what, that was worth it. That's actually a really easy take-home lesson in decision-making using AI.

SW: That's what I'm saying! But anyway, that's all I wanted to add. We can start the outro now.

IM: Okay, okay. I'll go again. (clears throat) ahem...

(outro music)

SW: If you like the show, give us a review on Apple Podcasts, Spotify, or wherever you are listening from. When you rate and review the show, you help us spread the word about Carry the Two so that other listeners can discover us.

IM: And for more on what's going on at IMSI, be sure to check us out online at our homepage: [IMSI dot institute](https://imsi.institute). We're also on BlueSky and on Instagram both at [IMSI dot institute](https://imsi.institute)! That's IMSI, spelled I M S I.

SW: Do you have a burning math question? Maybe you have an idea for a story on how mathematics and statistics connect with the world around us. Send us an email with your idea!

IM: You can send your feedback, ideas, and more to CarryTheTwo@imsi.institute.

SW: We'd like to thank Blue Dot Sessions for the music we use in Carry the Two.

IM: Peter Hall helped write the script for this episode of the podcast.

SW: Carry the Two is made possible by the Institute for Mathematical and Statistical Innovation, located on the gorgeous campus of the University of Chicago. We are supported by the US National Science Foundation and the University of Chicago.

IM: Until next time!

(outro)